

# Comparing odds ratios for nested subsets of dietary components

Martin Kulldorff,<sup>a</sup> Rashmi Sinha,<sup>b</sup> Wong-Ho Chow<sup>b</sup> and Nathaniel Rothman<sup>b</sup>

<b>Background</b>	In nutritional epidemiology, it is often of interest to disentangle the risk of disease associated with related foods or nutrients, where the food items are in a nested arrangement within a larger group. To compare odds ratios (OR) derived from a standard quantile-based analysis can be misleading since the amounts consumed may differ substantially for different dietary components.
<b>Methods</b>	The authors applied different logistic regression models on a case-control study concerning the risk of colorectal adenomas due to meat and its different subsets such as white meat, red meat and well-done red meat.
<b>Results</b>	By calculating OR for a fixed amount of intake, the authors suggest a method for partitioning the risk of one dietary item into that associated with increasingly detailed sub-components. A graph is presented for illustrating such partitions in terms of both addition and substitution effects.
<b>Conclusions</b>	Odds ratios based on upper versus lower quantiles or percentiles are useful as they compare the risk between the upper and lower ends of the consumption range. A complimentary set of OR are those based on fixed amounts of consumption. These allow for direct comparisons between nested subgroups of dietary components, in order to disentangle the risk linked to specific groups of foods or nutrients.
<b>Keywords</b>	Data interpretation, statistical, diet, epidemiological methods, nutrition assessment
<b>Accepted</b>	7 June 2000

In nutritional epidemiology, foods and nutrients are typically nested, so that some items under study are subsets of others. In such cases, it is often of interest to disentangle the risk of disease associated with the various nested subsets. Examples include the risk of colon cancer associated with consumption of different types of fibre,<sup>1</sup> fats,<sup>1,2</sup> or meats,<sup>1,3,4</sup> the risk of heart disease associated with alcohol from beer, wine and spirits,<sup>5,6</sup> and the risk of lung cancer associated with different vitamins such as  $\alpha$ - and  $\beta$ -carotene.<sup>7</sup>

When dissecting the risk of a dietary group into that which is generated by its various subsets, it is of interest to compare odds ratios (OR). For quantiles, such comparisons are misleading when the typical amounts consumed differ for different subsets. We illustrate this problem using an example of meat consumption, cooking practices and the risk of colorectal adenoma. An alternative approach allowing for direct comparison is to calculate all OR using the same fixed amount of intake, such as 10-g increments. We then demonstrate how to partition the

risk associated with total meat consumption to that of its sub-components.

## Materials and Methods

Our example uses data from a hospital-based case-control study of colorectal adenoma carried out at the US National Naval Medical Center. Described and analysed in detail elsewhere,<sup>8,9</sup> the study goal was to determine the relationship between meat cooking practices and the risk of colorectal adenoma. In short, subjects were 18 to 74 years old without a history of previous adenoma or a prior diagnosis of cancer except non-melanoma skin cancer. One-third were women. There were 146 incident cases, diagnosed with a histologically confirmed adenoma through either sigmoidoscopy or colonoscopy, and 228 controls, selected among patients without colorectal adenoma at sigmoidoscopy.

After returning home, participants filled in a self-administered food frequency questionnaire concerning their usual diet approximately one year before sigmoidoscopy/colonoscopy. Among other things they provided the frequency and portion size consumed of different types of meat and how well cooked they were. From that we calculated the estimated daily grams of meat consumed for each combination. A typical individual

<sup>a</sup> Division of Biostatistics, Department of Community Medicine, University of Connecticut School of Medicine, Farmington, CT 06030-6325, USA.

<sup>b</sup> Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA.

**Table 1** Odds ratios for the addition effect of different subsets of meats, using four nested models and three different methods of analysis based on quintiles, percentiles and fixed amounts of intake respectively.

	Categorized data					Continuous data	
	Q1	Q2	Q3	Q4	Q5	90th versus 10th percentile	10 g
<b>Model 1</b>							
Total meat	1.0	1.1	1.2	1.8	1.3	1.5	1.04
<b>Model 2</b>							
Red meat	1.0	1.0	1.6	1.8	2.5	2.3	1.11
White meat	1.0	1.0	0.7	1.0	0.8	0.6	0.95
<b>Model 3</b>							
Variably cooked red meat	1.0	0.5	1.1	1.3	1.6	2.0	1.17
Other red meat	1.0	1.7	1.1	1.5	2.0	1.2	1.04
White meat	1.0	0.9	0.7	1.0	0.8	0.7	0.95
<b>Model 4</b>							
Well done red meat	1.0	0.9	1.0	1.2	1.3	1.7	1.30
Medium/rare red meat	1.0	0.5	0.7	0.4	1.0	1.4	1.10
Other red meat	1.0	2.0	1.4	2.1	2.9	1.3	1.05
White meat	1.0	1.1	0.8	1.1	0.8	0.6	0.95

reported both well done and medium/rare consumption, but in varying quantities.

Odds ratios were calculated using logistic regression, with the meat variables being either in their original continuous form or categorized into quintiles. All analyses were adjusted for age, gender, total energy intake, physical activity, pack-years of smoking, non-steroidal anti-inflammatory drug use, and reason for colorectal screening.

Four sequentially nested models were analysed. In Model 1, total meat is the only meat variable included. In Model 2, this variable is divided into red meat (beef, pork and lamb) and white meat (chicken and fish). In Model 3, red meat is split into five types of red meat that are cooked to variable levels (beef steak, hamburgers/cheeseburgers, pork chops/ham steaks, sausage/hot dogs and bacon), referred to as *variably cooked red meat*. Also in this model are the remaining types of red meat that are either cooked to a standard level or for which it was not known how well they were cooked (liverwurst, liver, luncheon meats, beef stew, pork roast, spaghetti sauce with meat, other ground beef, and lamb) are referred to as *other red meat*. Finally, in Model 4, the variably cooked red meat is split into whether it was prepared well done or medium/rare. White meat was retained in Models 3 and 4 and 'other red meat' was retained in Model 4, so that total meat consumption was accounted for in all models.

We looked at addition effects, the risk associated with adding the amount of one subgroup (e.g. red meat) while keeping the consumption of other subgroups (e.g. white meat) constant, as well as substitution effects, where the total consumed is held fixed so that adding to the amount consumed of one subgroup (red meat) is offset by subtracting an equal amount consumed of the other subgroups (white meat). The OR of the substitution effect is 1 if there is no difference in the risk associated with red and white meat. By including red and white meat as two separate variables in the regression equation we obtained estimates and confidence interval (CI) for the addition effects. When instead the total meat variable is used, this reflects the

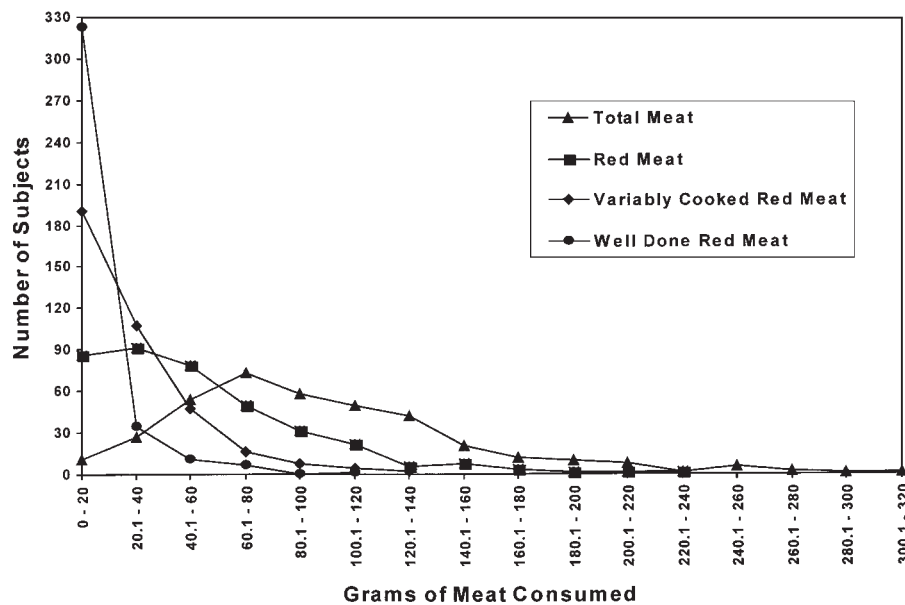
addition effect of white meat (red meat is fixed) while the red meat variable provides the estimate and CI for the substitution effect of red versus white meat (total meat is fixed).<sup>10</sup>

## Results

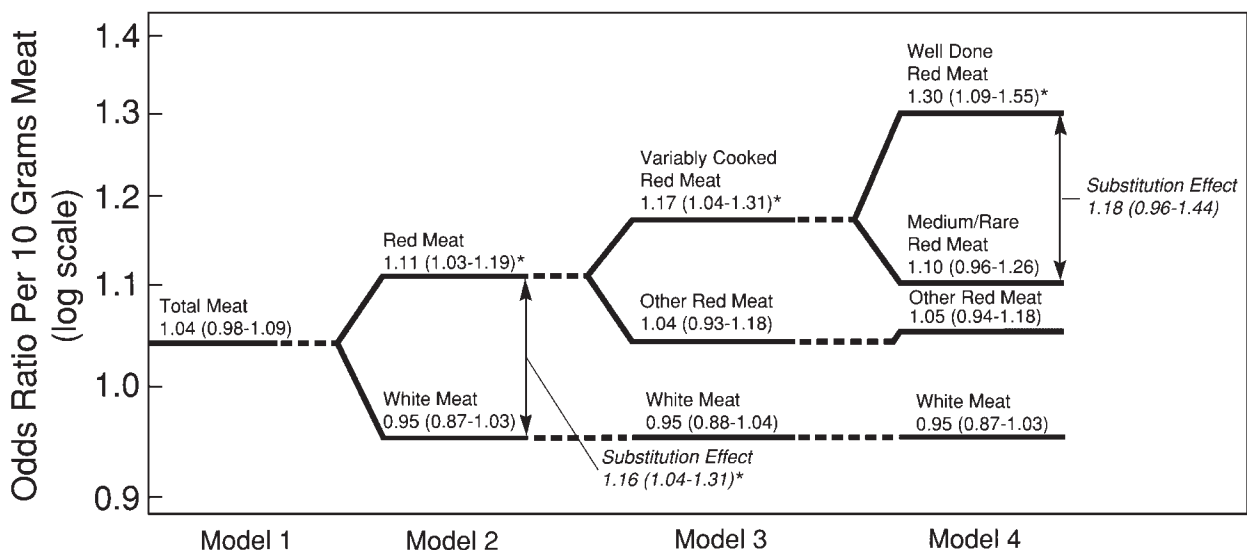
The study hypothesis was that there is an increased risk of colorectal adenoma with meat intake, in particular with red meat, and especially with well done red meat,<sup>3,4</sup> as these subsets contain increasing quantities of heterocyclic amines, known carcinogens in animals.<sup>11</sup> With a traditional analysis using data categorized into quintiles, the OR for the fifth versus first quintile increases if we compare total meat in Model 1 (OR = 1.3) with red meat in Model 2 (OR = 2.5), as expected based on the hypothesis (Table 1). However, the OR is lower for variably cooked red meat in Model 3 (OR = 1.6) and even lower for well done red meat in Model 4 (OR = 1.3). Similar results are obtained for continuous data when comparing the OR for the 90th and 10th percentiles (i.e. the medians of the 5th and 1st quintiles) with the OR being 1.5, 2.3, 2.0 and 1.7, respectively.

At first glance, this may give the impression that well done red meat is associated with a lower risk than red meat *per se*. However, an OR must of course always be interpreted in terms of the difference in consumption levels compared. Figure 1 shows the distribution consumed of different subsets of meat. The differences in consumption between the 90th and 10th percentiles are 125 g for total meat, 92 g for red meat, 56 g for variably cooked red meat and 24 g for well done red meat. So, when using quintiles, we are comparing 'apples' with 'oranges', and given the hypothesis, it is not surprising that there is a higher OR for a 92-g difference in red meat consumption than for a 24-g difference in well done red meat consumption.

To put the OR on an equal footing for comparison, it is necessary to calculate them using the same fixed amount of intake. In the last column of Table 1 this is done for a 10-g increment in average daily consumption. The estimated OR increases from total meat (OR = 1.04, i.e. a 10-g increment in



**Figure 1** The empirical distribution of the consumption of total meat, red meat, variably cooked red meat, and well done red meat



**Figure 2** Graphical representation showing how the risk associated with total meat consumption is partitioned into the risks associated with various sub-components. Odds ratios (OR) for 10-g increments in consumption, together with 95% CI, are given for both addition and substitution effects. On the log scale, the addition effect is the distance between the thick line for a particular meat type and an imaginary line corresponding to an OR of 1, while the substitution effect is the distance between two thick lines representing two different meat types in the same model

consumption increases the odds by an estimated 4%), to red meat (OR = 1.11), to variably cooked red meat (OR = 1.17) to well done red meat (OR = 1.30), consistent with the prior hypothesis.

A way to depict the partition of OR for a nested set of variables is through a graph as illustrated in Figure 2. Marked on the left side is the OR of 1.04 associated with a 10-g increase in total daily meat consumption. As we move right in the Figure, this is then partitioned into an OR of 1.11 for red meat and 0.95 for white meat, and so on. As another example, the

OR for variably cooked red meat is 1.17, which is partitioned into 1.10 for medium/rare and 1.30 for well done red meat.

All OR in Table 1 represent addition effects. Figure 2 presents a natural illustration of the relationship between all the different addition and substitution effects within a nested group of models. The y-axis reflects the OR for the addition effects of different subgroups of meat as taken from Table 1. When the y-axis is drawn on the log scale, as in Figure 2, the log substitution effect for two different subgroups is simply the distance between their log OR as illustrated in the Figure. For example,

**Table 2** With a quadratic term for total meat, these are odds ratios (OR) for the addition effect of 10 g of different types of meats, when the total meat consumption is 0, 10, 20, 50 and 100 g per day, respectively, using four nested models. The addition effect depends on the level of total meat consumed because of the quadratic term, but irrespective of level, the effect can be dissected into its respective components for different subsets of meat by looking down the same column. In the rightmost column are the effects of substituting 10 g of one type of meat for 10 g of another type, comparing red versus white meat in Model 2, variably cooked versus other red meat in Model 3, and well done versus medium/rare red meat in Model 4. Even though there is a quadratic term for total meat, these substitution effects do not depend on the total amount of meat consumed

	Addition effect of 10 g when total meat consumption is					Substitution effect of 10 g
	0g	10g	20g	50g	100g	
<b>Model 1</b>						
Total meat	1.024	1.025	1.026	1.029	1.033	–
<b>Model 2</b>						
Red meat	1.114	1.114	1.113	1.111	1.107	1.16
White meat	0.957	0.957	0.956	0.954	0.951	–
<b>Model 3</b>						
Variably cooked red meat	1.177	1.177	1.176	1.173	1.168	1.12
Other red meat	1.055	1.054	1.053	1.051	1.047	–
White meat	0.960	0.960	0.959	0.957	0.953	–
<b>Model 4</b>						
Well done red meat	1.306	1.305	1.304	1.302	1.299	1.18
Medium/rare red meat	1.109	1.108	1.107	1.106	1.103	–
Other red meat	1.060	1.059	1.058	1.057	1.054	–
White meat	0.954	0.954	0.953	0.952	0.949	–

the substitution effect of consuming well done rather than medium/rare red meat is  $1.30/1.10 = 1.18$ , so on the natural log scale we have that  $\ln(1.18) = \ln(1.30) - \ln(1.10)$ .

In the example above we used a model where the risk associated with increased consumption changes according to a linear function on the log scale. That fits very well with the data, giving a higher log likelihood value than the categorical models. Adding a quadratic term did not significantly improve the model fit. This may or may not be true for other data set though, and the proposed dissection and comparisons can also be used when the regression equation contains non-linear terms, if these non-linear terms represent the total consumed of the different sub-components.

Using the same data, Table 2 presents the dissection of the OR when a quadratic term is added to the model. The nature of a non-linear term is that the increase in the odds due to a fixed increase in consumption is different depending on the original amount consumed. For example, when fitting a quadratic rather than linear term, the OR for a 10-g increase of total meat are 1.024, 1.025, 1.026, 1.029 and 1.033 for original consumptions of 0, 10, 20, 50 and 100 g, respectively. As before, it is possible to dissect how much of this risk is due to various sub-components. This is done by including in the logistic regression model a linear term for each of the meat subtypes, plus a quadratic term for total meat. For example, with our data the OR of 1.026 associated with an increase of total meat consumption from 20 to 30 g is dissected into OR of 1.113 and 0.956 for red and white meat, respectively.

Table 2 should be compared with the rightmost column of Table 1, and it can be seen that the OR are very similar. This is expected since the quadratic term was not significant for this data. There are differences in the OR depending on the total meat consumed, as for example, the OR is 1.306 for a 10-g increase of well done red meat if the original total meat consumption is 0 g, while the OR is 1.299 if the original total

consumption is 100 g. Likewise, the OR for medium/rare red meat are 1.109 and 1.103, respectively. These differences would be bigger in other data sets where the non-linear term was significant. The relative size of the OR of different sizes do not change though, and the substitution effect does not depend on the total amount consumed, even when there is a non-linear term in the model for total meat consumption. For example, the OR for substituting 10 g of well done for 10 g of medium/rare red meat is  $e^{\ln(1.306) - \ln(1.109)} = e^{\ln(1.299) - \ln(1.103)} = 1.18$ .

## Discussion

Whether the data are categorized or analysed in their continuous form, OR for different nested subsets of dietary components cannot be compared when based on quantiles or percentiles. In order to compare risks for 10-g increments, it is necessary to analyse continuous data. Dorfman *et al.*<sup>12</sup> looked at different models for epidemiological studies containing an exposure variable that is classified by type, including models containing ratios of related variables. They concluded that using a continuous scale variable for each subset is preferable compared to either ratios or categorization. The ability to compare and partition the OR, as suggested in this paper, is an added argument for that. When using continuous variables though, it is important not to extend the inference outside the interval of observations. For example, since very few subjects in our data set ate more than 100 g daily of well done red meat, and none more than 140 g, we cannot estimate the effect of eating 100 additional daily grams of such meat nor the effect of eating 10 additional daily grams for a person whose original consumption is 150.

In our example, there are a total of 10 different substitution effects that can be calculated, one in Model 2, three in Model 3 and six in Model 4, as every dietary component can be substituted for every other component in the same model. The OR for each of these substitution effects are easily read using Figure 2,

by measuring the distance between the two dietary components that are to be substituted, and comparing that distance with the scale on the y-axis. A table reporting 10 additional and 10 substitution effects would, by contrast, be much harder to comprehend.

When applying and interpreting the proposed type of analysis, some caution is warranted. It is inadvisable to split a variable into two components that are highly correlated with each other. How correlated they can be depends on the size of the data set. The key way to determine whether a dissection is informative is to look at the width of the CI for both the addition and substitution effects. It is also important to note that the proposed method cannot be used to compare OR for exposures that are measured in different units. That is a different type of problem, related to attributable risk.<sup>13</sup>

Odds ratios based on upper versus lower quartiles or quintiles are useful, as they compare the risk between the upper and lower ends of the consumption range in the population under study. Hence the OR reflect a risk differential that is reasonably achievable through preventive measures. A complimentary set of OR are those based on fixed amounts of consumption. These allow for direct comparisons between nested subgroups of dietary components, in order to disentangle the risk linked to specific groups of foods or nutrients. They may also be applied for other types of exposure that have natural sub-components, such as physical activity, occupational histories, radiation exposure or smoking.

## References

- <sup>1</sup> Willett WC, Stampfer MJ, Colditz GA, Rosner BA, Speizer FE. Relation of meat, fat, and fiber intake to the risk of colon cancer in a prospective study among women. *N Engl J Med* 1990;**323**:1664–72.
- <sup>2</sup> Howe GR, Aronson KJ, Benito E *et al.* The relationship between dietary fat intake and risk of colorectal cancer: evidence from the combined analysis of 13 case-control studies. *Cancer Causes Control* 1997;**8**:215–28.
- <sup>3</sup> Gerhardsson de Verdier M, Hagman U, Peters RK, Steineck G, Övervik E. Meat, cooking methods and colorectal cancer: a case-referent study in Stockholm. *Int J Cancer* 1991;**49**:520–25.
- <sup>4</sup> Schiffman MH, Felton JS. Fried foods and the risk of colon cancer. *Am J Epidemiol* 1990;**131**:376–78.
- <sup>5</sup> Nanji AA. Alcohol and ischemic heart disease: wine, beer or both? *Int J Cardiol* 1985;**8**:487–89.
- <sup>6</sup> Rimm EB, Klatsky A, Grobbee D, Stampfer MJ. Review of moderate alcohol consumption and reduced risk of coronary heart disease: is the effect due to beer, wine or spirits? *Br Med J* 1996;**312**:731–36.
- <sup>7</sup> Ziegler RG, Colavito EA, Hartge P *et al.* Importance of  $\alpha$ -carotene,  $\beta$ -carotene, and other phytochemicals in the etiology of lung cancer. *J Nat Cancer Inst* 1996;**88**:612–15.
- <sup>8</sup> Sinha R, Chow WH, Kulldorff M *et al.* Well done, grilled red meat, 1-amino-3,8-dimethylimidazo[4,5-f]quinoxaline (MeIQx), methabolizing enzymes and risk of colorectal adenomas. *Proceedings of the 88th Annual Meeting of the American Association for Cancer Research* 1998; **39**:365–66.
- <sup>9</sup> Sinha R, Chow WH, Kulldorff M *et al.* Well done, grilled red meat increases the risk of colorectal adenomas. *Cancer Res* 1999;**59**:4320–24.
- <sup>10</sup> Kipnis V, Freedman LS, Brown CC, Hartman AM, Schatzkin A, Wacholder S. Interpretation of energy adjustment models for nutritional epidemiology. *Am J Epidemiol* 1993;**137**:1376–80.
- <sup>11</sup> Sugimura T, Sato S, Wakabayashi K. Mutagens/carcinogens in pyrolysate of amino acids and proteins in cooked foods: heterocyclic aromatic amines. In: Woo YT, Lai DY, Arcos JC, Argue MF (eds). *Chemical Induction of Cancer, Structural Bases and Biological Mechanisms*. New York: Academic Press, 1988, pp.681–710.
- <sup>12</sup> Dorfman A, Kimball AW, Friedman LA. Regression modeling of consumption or exposure variables classified by type. *Am J Epidemiol* 1985;**122**:1096–107.
- <sup>13</sup> Lee WC. Characterizing exposure-disease association in human populations using the Lorenz curve and Gini index. *Stat Med* 1997; **16**:729–39.